

EXHIBIT 159

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 204

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

RICHARD KADREY, ET AL.,)
INDIVIDUAL AND REPRESENTATIVE)
PLAINTIFFS,) LEAD CASE NO.
v.) 3:23-cv-03417-VC
META PLATFORMS, INC.,)
DEFENDANT.)

)

* * * H I G H L Y C O N F I D E N T I A L * * *
* * * A T T O R N E Y S ' E Y E S O N L Y * * *

VIDEO-RECORDED 30(B) (6) DEPOSITION OF

MICHAEL CLARK

VOLUME III

WEDNESDAY, NOVEMBER 20, 2024

SAN FRANCISCO, CALIFORNIA

9:38 A.M. PST

REPORTED BY AUDRA E. CRAMER, CSR NO. 9901

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 293

1 BY MS. POUEYMIROU:

2 Q. So this document, this page 89800,
3 shows Meta removing copyright information,
4 including data publication, author, ISBN
5 numbers, disclaimers about not selling, sharing
6 or giving way because it is an infringement.

7 And Meta achieved this via a script
8 that Mr. Bashlykov created?

9 MS. HARTNETT: Object to the form.

10 THE WITNESS: Mr. Bashlykov wrote a
11 script with the logic that is outlined on Bates
12 page 89798 and continues on 89799 to remove and
13 clean up text that matched this description,
14 which would include the items that you spoke
15 about.

16 BY MS. POUEYMIROU:

17 Q. And does Meta still have this script?

18 A. I do not know. I walked through with
19 Mr. Bashlykov what the logic of the script was.
20 I did not look at the actual script.

21 Q. Do you know whether this script was
22 used with Books3?

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 294

1 MS. HARTNETT: Object to the form.

2 THE WITNESS: This script, in talking
3 to Mr. Bashlykov, was specific to LibGen.

4 BY MS. POUEYMIROU:

5 Q. Do you know whether Books3 also had its
6 copyright information stripped from it?7 MS. HARTNETT: Objection to the form of
8 the question.

9 THE WITNESS: I do not know.

10 BY MS. POUEYMIROU:

11 Q. Do you know whether Meta used this
12 script on Project Gutenberg?13 A. Project Gutenberg was processed at the
14 same time as Books3. This script was written
15 specific to LibGen and to this criteria then,
16 which would postdate the processing of
17 Gutenberg.18 Q. And who told Mr. Bashlykov to create
19 this script?

20 MS. HARTNETT: Object to the form.

21 THE WITNESS: The logic in this script
22 is part of making the data usable and looking

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 295

1 for traits of the data that make the data usable
2 for training. And so as part of the cleanup
3 that they were doing in the data to get ready
4 for the data pipeline, they wrote this in order
5 to make the data usable and not have repetitive
6 patterns and not have extra line spaces and not
7 have things that would decrease the quality of
8 the model.

9 BY MS. POUEYMIROU:

10 Q. By removing the copyright information
11 from the training data, that means that there is
12 no concern that -- a memorization concern or a
13 regurgitation concern; is that correct?

14 MS. HARTNETT: Objection to the form
15 and to the vagueness of "copyright information."

16 THE WITNESS: Could you be more
17 specific?

18 BY MS. POUEYMIROU:

19 Q. Yes. So memorization, as we've
20 discussed, you have the regurgitation of
21 training data.

22 If you haven't trained on copyright

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 296

1 information, there's no risk of regurgitating
2 that information; is that correct?

3 MS. HARTNETT: Objection to the form.

4 THE WITNESS: Is that a hypothetical,
5 or is that specific --

6 BY MS. POUYEMIROU:

7 Q. No. I'm asking: Can you regurgitate
8 data that you have not trained on?

9 MS. HARTNETT: Objection to the form.

10 THE WITNESS: Potentially. Like, any
11 phrase could potentially be -- could you be more
12 specific? I'm --

13 BY MS. POUYEMIROU:

14 Q. So Meta removed files that
15 self-identified as pirated from its training
16 data; is that correct?

17 A. As of the training time of LLaMA 2,
18 items that were self-identified as pirated or
19 stolen were removed, yes.

20 Q. And when we discussed that last week,
21 the removal of pirated material was also
22 connected to a concern that it would be

11/20/2024

Richard Kadrey, et al. v. Meta Platforms, Inc. Michael Clark 30(b)(6), Vol III
Highly Confidential - Attorneys' Eyes Only

Page 461

1 STATE OF CALIFORNIA)

2 COUNTY OF LOS ANGELES) SS.

3 I, AUDRA E. CRAMER, CSR No. 9901, in and for the

4 State of California, do hereby certify:

5 That, prior to being examined, the witness named

6 in the foregoing deposition was by me duly sworn to

7 testify the truth, the whole truth and nothing but the

8 truth;

9 That said deposition was taken down by me in

10 shorthand at the time and place therein named, and

11 thereafter reduced to typewriting under my direction,

12 and the same is a true, correct and complete transcript

13 of said proceedings;

14 I further certify that I am not interested in the

15 event of the action.

16 Witness my hand this 22 day of November,

17 2024.

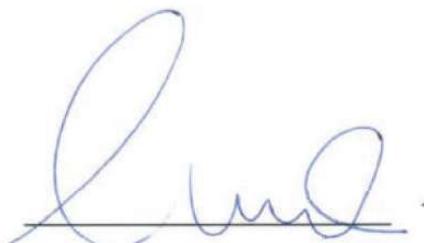
18

19

20

21

22


Certified ShorthandReporter for the
State of California